

59-82

121956

P-19

N98-15046

**MASS STORAGE SYSTEM
EXPERIENCES AND FUTURE NEEDS
AT
THE NATIONAL CENTER FOR ATMOSPHERIC
RESEARCH**

**Summary of the Presentation to the Conference on
Mass Storage Systems and Technologies for
Space and Earth Science Applications
July 23-25, 1991**

b y

**Bernard T. O'Lear
Manager, Systems Programming
Scientific Computing Division
National Center for Atmospheric Research**

3/5/91

This is a summary of the presentation given at the Conference on Mass Storage Systems and Technologies for Space and Earth Science Applications. The presentation was compiled at the National Center for Atmospheric Research (NCAR), Boulder, Colorado. NCAR is operated by the University Corporation for Atmospheric Research and is sponsored by the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author and do not necessarily reflect the views of the National Science Foundation.

This presentation is designed to relate some of the experiences of the Scientific Computing Division at NCAR dealing with the "data problem." A brief history and a development of some basic Mass Storage System (MSS) principles are given. An attempt is made to show how these principles apply to the integration of various components into NCAR's MSS. There is discussion of future MSS needs for future computing environments.

NCAR provides supercomputing and data processing for atmospheric, oceanic and related sciences. This service is provided for university scientists and for scientists located at NCAR. There is a total of about 1200 users.

The data problem for this community can briefly be summarized as follows; Historical atmospheric data is archived, programs are saved and the data which model the atmosphere, oceans and sun are saved. The NCAR storage experience is based upon current supercomputing megaflop rates which produce a number of terabytes archived on a yearly basis. There is a history of data growth and file growth. The NCAR data storage experience has been as follows; There are about 500 bytes of information archived for each megaflop of computing. When NCAR had an X-MP/48, the archive rate for the utilized megaflop compute rate was 3 terabytes per year. The installation of a Y-MP8/864 increased the archival rate to 6 terabytes per year. Forecasting future computing configurations and atmospheric models being planned we are now approximating a 30-50 terabyte archive per year rate by the year 1993 or 1994.

Data has been saved in many forms over NCAR's existence and then migrated to machine-readable media. Some of the data has come from handwritten logs, from punch cards, half-inch tape. All of this has been collected and is now archived on IBM 3480 cartridge tape. One of the basic principles for archiving this data is to identify certain classes of data. Archive data is kept forever. Long-term

data is kept for 10 to 15 years. Near-term data is kept for 1 month to 1 year and a category called scratch data is killed after 1 month and cannot be recovered automatically by the system.

One of the other basic principles that has been identified is that dataset sizes continue to grow as a function of supercomputing sizing. The amount of data that can be saved is bound in storage by media capacities. That is, these criteria are established for determining which data will be saved and for how long because there is not an infinite media capacity at this time. Our experience has shown that every 10 to 15 years the data in the MSS will need to be migrated to a new media base because of changing systems and obsolescence of existing media. Usually the media or the drives cannot be purchased anymore. This migration takes place not because the data is bad on the media, but because the drives will not be available.

Another problem is that a number of companies have provided the capability for this massive storage, but the small companies tend to disappear within five years. The drive components that have been furnished for mass data storage disappear in five to eight years no matter what company they come from.

The next basic principle is that the migration of the mass storage system data to a new media base, which is now several ten's of terabytes, is not a trivial operation. The migration does not take place in a short amount of time. For instance, one-time migrations can run for long periods of time, necessarily years to move terabytes data. It is very difficult to guarantee that the data is migrated absolutely without reading it back, which is time consuming. These migrations are very costly and in my opinion shouldn't be done. We have developed the concept of "DATA OOZE," and we prefer this technique over migration right now. The way DATA OOZE works is that it is a continuous movement of data within the system. The data is moving across the

storage hierarchy and across the changing media types under the control of the MSS. The migration path for this data in the hierarchy can be from memory to solid state disk to high speed disk to disk arrays or farms, and from there out to some kind of tape. Later on as new data storage media become available, the data is migrated onto these media in real time, since every day some amount of the data is migrated as it is being used.

Our conclusions from these experiences have been that new components and media types are integrated according to the following rules; Use standard components. The standards may be real or de facto and apply in the areas of channels, interfaces, operating systems, media, etc. We look for media that is easy to obtain and is cost effective. We look for the long-term viability of the vendor and multiple sources for the many system components. In the area of mass storage system integration we look at access speeds, ease of expandability, heterogeneous host access, maintenance costs, media costs and systems costs.

There are a number of future growth issues for the NCAR MSS. The Scientific Computing Division (SCD) continues to develop future configuration scenarios. These scenarios try to anticipate the functional requirements we anticipate providing for our scientific community. There are three key components we need to address: network services and access, the large scale computing (Big Iron), and the data archives. Of course, these all play within the context of distributed computing.

The near-term issues for the NCAR MSS focus on some immediate upgrades which will deal with the MSS growth for a couple of years. The entire archive will be migrated onto double density 3490 and 3490-compatible media. The mid-90s to late 90s became more interesting because of the expanding interest in archiving vast data collections.

The issues of future growth will be centered in three areas of ongoing development: the various MSS software packages, the data storage components and the networks.

The questions then become how all of these components get assembled and which ones do we plan to use. Will SCD be able to construct an effective peta-byte MSS by the end of the decade? Which of our basic principles can we apply to insure that such a system can be built?

###

**Mass Storage System
Experiences and Future Needs
at
The National Center for Atmospheric Research**

**Conference on Mass Storage Systems
and Technologies for
Space and Earth Science Applications**

July 23-25, 1991

**Bernard T. O'Lear
Manager, Systems Programming
Scientific Computing Division
National Center for Atmospheric Research**

 **NCAR Scientific Computing Division**
Supercomputing • Communications • Data

**The following presentation was compiled at the
National Center for Atmospheric Research
Boulder, Colorado**

- **The National Center for Atmospheric Research is operated by the University Corporation for Atmospheric Research and is sponsored by the National Science Foundation.**
- **Any opinions, findings, conclusions, or recommendations expressed in this talk are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.**

 **NCAR Scientific Computing Division**
Supercomputing • Communications • Data

Introduction

**The experiences of a scientific center dealing with
"The Data Problem"**

- Brief history
- The current computing environment
- Development of some basic principles
- How the principles apply
- Future needs for future computing environments



NCAR Scientific Computing Division
Supercomputing • Communications • Data

History

**NCAR provides supercomputing and data processing for atmospheric,
oceanic, and related sciences:**

- At universities
- At NCAR
 - Totals about 1200 users

The data problem for this community

- Save and archive historical atmospheric data
- Save programs and data which model the atmosphere, oceans, and sun



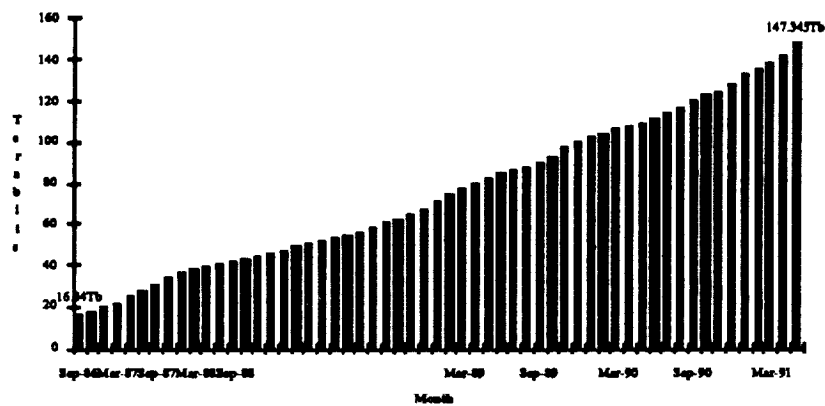
NCAR Scientific Computing Division
Supercomputing • Communications • Data

The NCAR Storage Experience

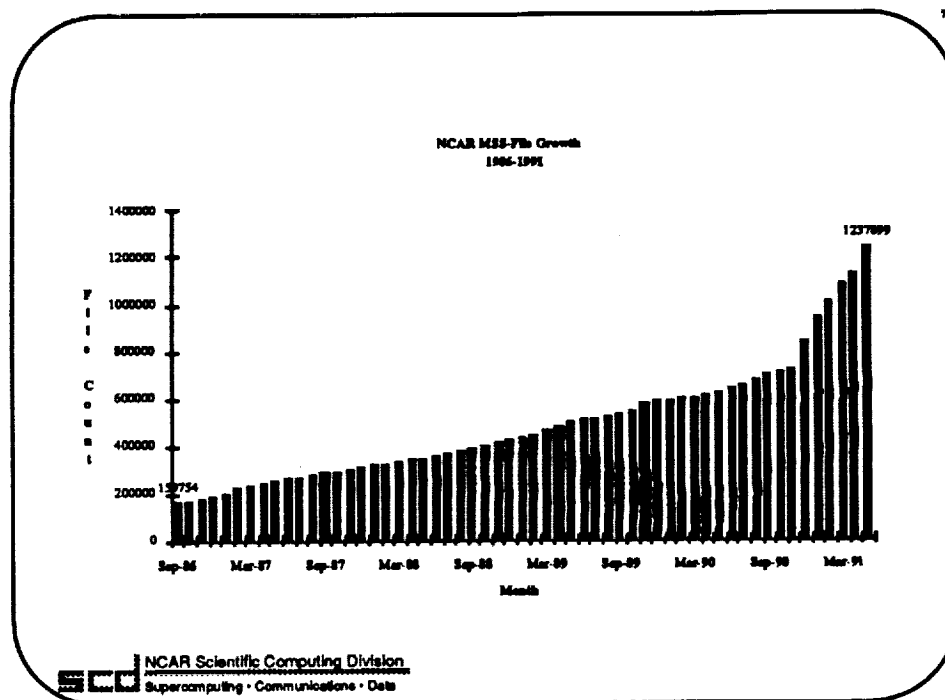
- 500 Bytes per million flop
- Archival rate for model output
 - 4 TBytes/year with X-MP/48
 - 8 TBytes/year with Y-MP8/864
 - 40 TBytes for climate simulation

SCD NCAR Scientific Computing Division
Supercomputing • Communications • Data

NCAR MSS Data Growth
1986-1991



SCD NCAR Scientific Computing Division
Supercomputing • Communications • Data



NCAR Mass Storage Systems (MSS)

Usage Data

- 101,700 tape cartridges in use
- Over 18.5 Tbytes of data stored
- Over 710,000 files
- Average file length 26.2 MB

Fast Path

≤ 2 minute delivery

NCAR Scientific Computing Division
Supercomputing • Communications • Data

History (Continued)

Data saved in many forms -- then migrated to machine readable media:

- Handwritten logs -----> Punched cards
- Punched cards -----> One-half inch tape
- One-half inch tape -----> AMPEX TBM tape
- AMPEX TBM tape -----> IBM 3480 tape
- IBM 3480 tape -----> IBM 3490-E tape
- IBM 3490-E tape -----> ? ? ?

Basic Principles

Identification of data classes:

- Archive data = keep forever
- Long-term data = keep 10-15 years
- Near-term data = keep 1 month to 1 year
- Scratch data = kill after 1 month

Basic Principles (Continued)

- Dataset sizes continue to grow as a function of supercomputer sizing
- Dataset sizes are constrained in Storage by media capacities
- Every ten to fifteen years, the data in the MSS will need to be migrated to a new media base

Basic Principles (Continued)

Migration:

- Not because the data is bad on the media...
- But because the drives will not be there
 - Half life of a start-up company = 5 years
 - Half life of drive electronics = 5-8 years

Basic Principles (Continued)

The migration of MSS contents ("n" tera-bytes) to new media is not a trivial operation

One-time migrations:

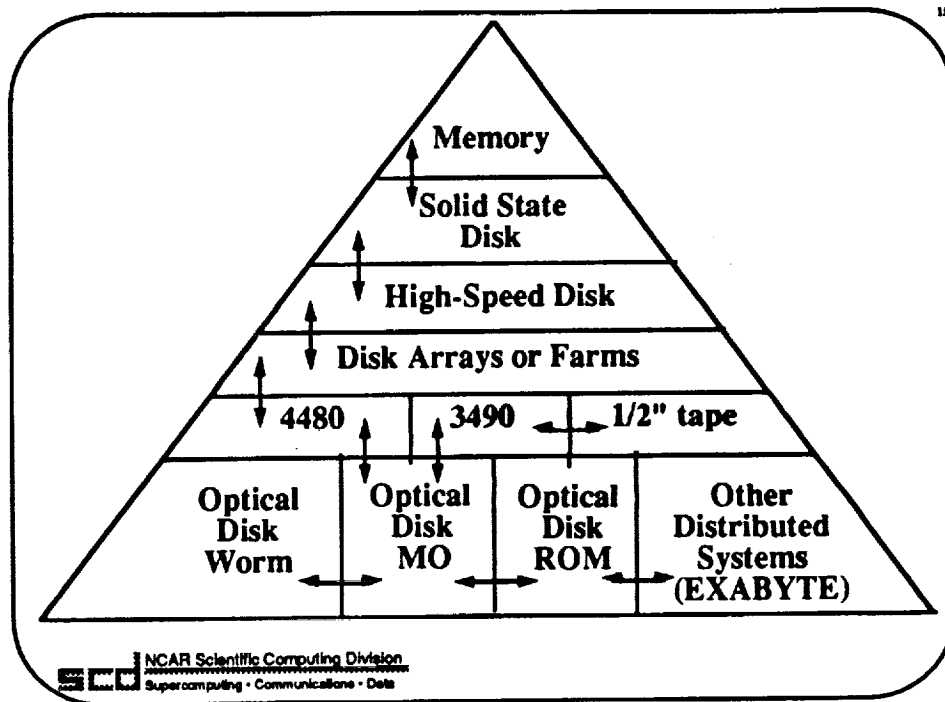
- Run for long periods of time (years)
- Are difficult to guarantee
- Are costly
- Shouldn't be done

Basic Principles (Continued)

Data OOZE preferred over migration

Data OOZE is a continuous movement of data within the system:

- Data movement across:
 - The storage hierarchy
 - The changing media types



14

CONCLUSIONS

New components and media types are integrated according to these rules:

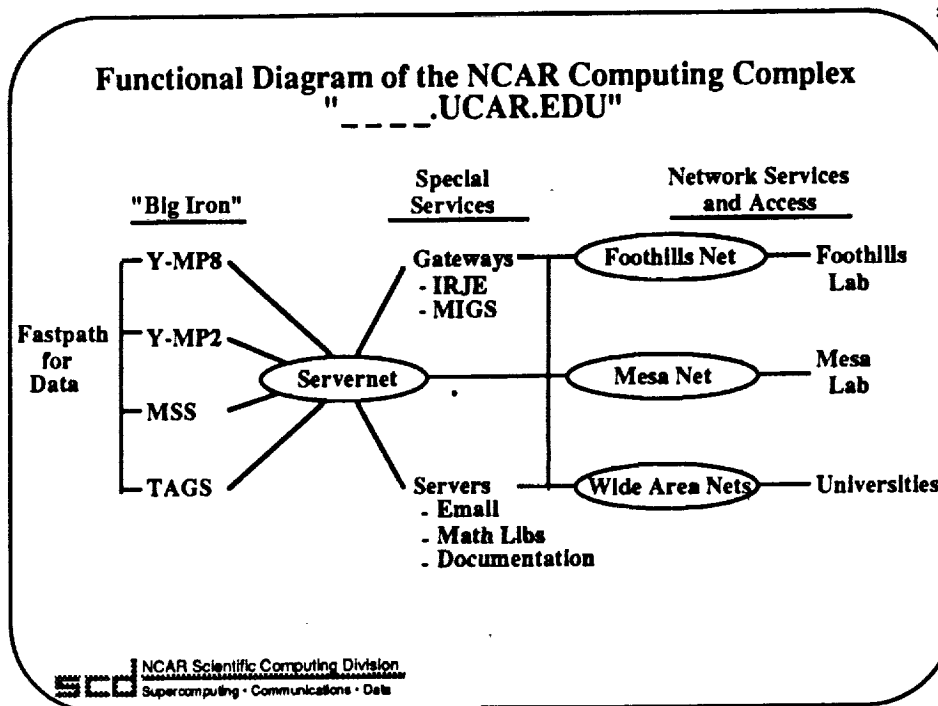
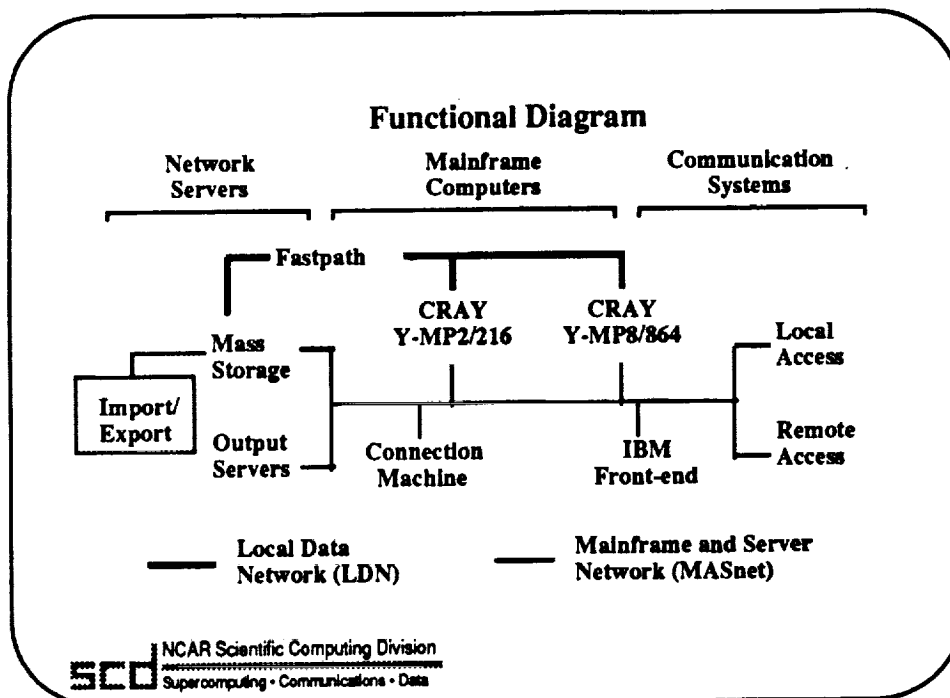
- Standards (real or de facto)
 - For channels and interfaces (IBM, IPI, HIPPI, SCSI)
 - For media
- Long-term viability of vendor
- Multiple source availability for media (drives?)

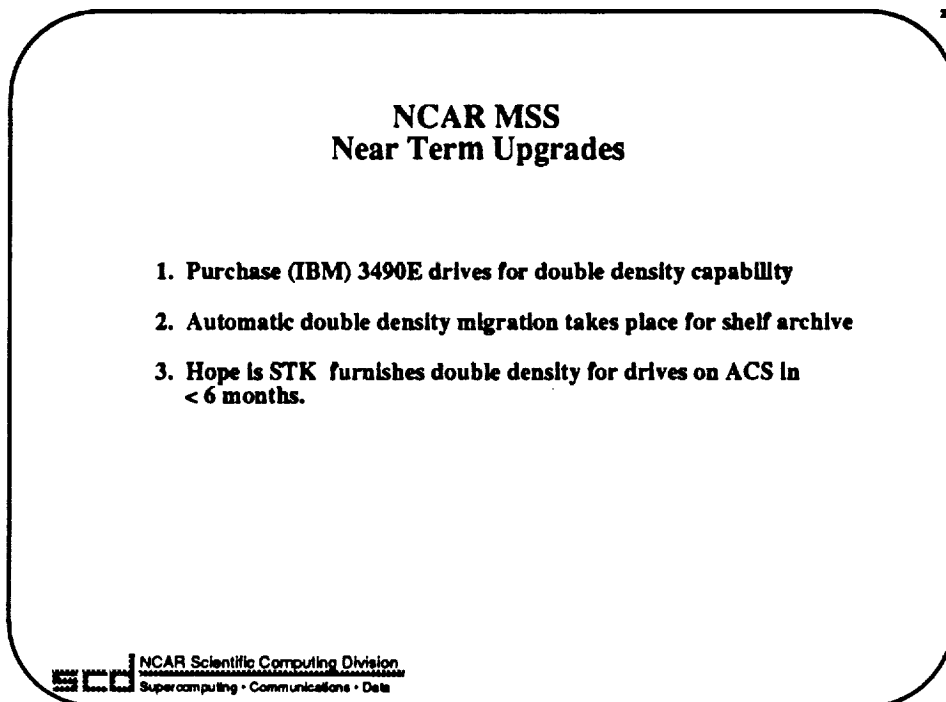
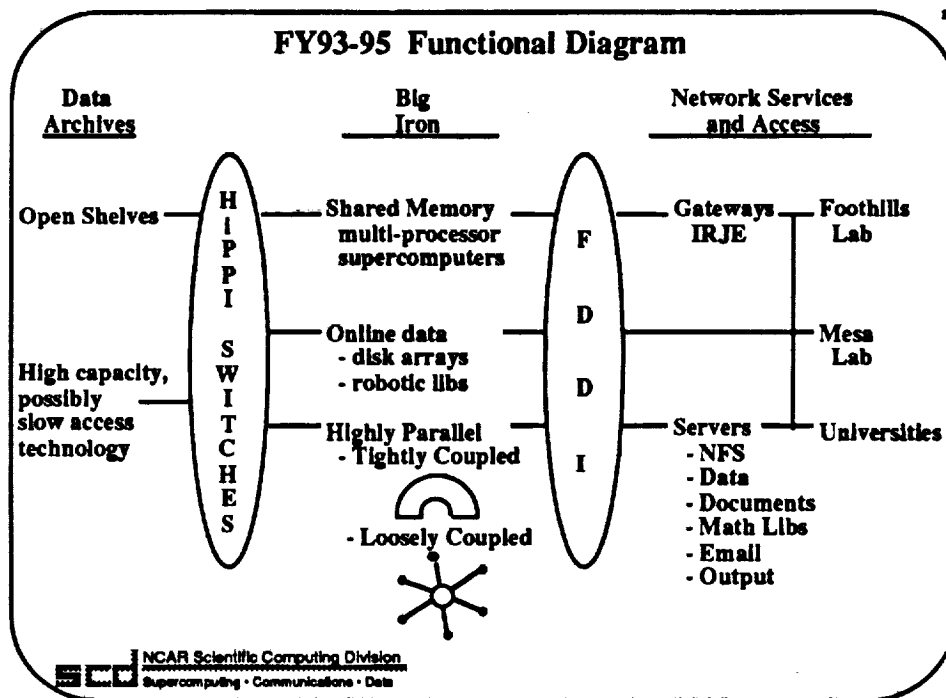
NCAR Scientific Computing Division
Supercomputing • Communications • Data

MSS Integration

- Access speeds (sometimes)
- Ease of expandability
- Multiple heterogeneous host access
- Maintenance costs
- Media costs
- System cost

FUTURE GROWTH ISSUES IN THE NCAR MASS STORAGE SYSTEM





NCAR MSS

The Issues of Future Growth are dependent upon:

1. Future MSS Software


- a. Distributed MSS
- b. Large archives (Peta-Byte)

2. Future Data Storage

- a. The media
- b. The drives
- c. The robotics

3. The Network and Channels

- a. HIPPI
- b. Fibre channel standards
- c. fabric (switch)

 NCAR Scientific Computing Division
Supercomputing • Communications • Data

1. Future MSS Software

a. Distributed MSS

- UNITREE (DISCOS)
- Infinite Storage Architecture (EPOCH)
- Distributed Physical Volume Repository (EPOCH & STK)
- EMAS (E-SYSTEMS)
- NASTore (NASA, Ames)
- NETARC and AWBUS (CDC)
- SWIFT (IBM)
- DataMesh (Hewlett Packard)
- M (DS) ²NASA Goddard

b. Peta-Byte Archives

- How do we build them?

 NCAR Scientific Computing Division
Supercomputing • Communications • Data

2. Future Data Storage

a. THE MEDIA

- The 3M National Media Laboratory (Media Database)
- Government funds and goals
- Private sector participation
- Standards being developed

b. THE DRIVES

- Being developed for the media
- 10-year life span
- Attachable to various robotics

c. THE ROBOTICS

- StorageTek is the leader
- ODETICS
- EXABYTE and others

3. The Network and Channels

a. Standards moving fast for HIPPI

- The HIPPI switch

b. Fibre Channel Advantages

- Length to 10 kilometers
- General Protocol
 - HIPPI
 - SCSI
 - IPI
 - Others
- Security
- Immune to Electrical Disturbance

c. Fabric Switch

